DOCUMENT RESUME

ED 236 217 TM 830 737

AUTHOR Maloney, Eileen; And Others
TITLE A Validity Study of an Editing Cloze Proficiency Test for Students of English as a Second Language.
PUB DATE 13 Apr 83
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; College Entrance Examinations; *English (Second Language); *Foreign Students; Higher Education; *Language Tests; Malay; Second Language Learning; *Test Validity
IDENTIFIERS Center for English Second Language Placement Test; *Davies Speeded Reading Test English Language Prof; Michigan Test of Aural Comprehension; Michigan Test of English Language Proficiency

ABSTRACT

Since the numbers of foreign students aspiring to college and graduate degrees in the United States is increasing, there is a need for a reliable and valid English proficiency test which, ideally, would be neither costly nor time consuming. The purpose of this study was twofold. First, the extent to which three tests, the Michigan Test of English Language Proficiency and Aural Comprehension, the Center for English as a Second Language Placement Test, and the Davies Speeded Reading Test of English Language Proficiency (an editing cloze test), tended to measure the same underlying construct was explored. Second, two methods of grading the Davies, the Omission method and the Omission and Commission method, were compared. A total of almost 300 subjects in four groups were tested. Two were homogeneous groups of Malaysians relatively proficient in English. Two consisted of foreign students with diverse language backgrounds. One of these groups had already completed an intensive proficiency course in English, while the other group included students at all levels of English language proficiency. Reliability and validity coefficients were reported along with test means and standard deviations. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236217

A Validity Study of an Editing Cloze Proficiency Test for Students of English as a Second Language

Eileen Maloney

Center for English as a Second Language
Southern Illinois University at Carbondale

Susan M. Tracz

Guidance & Educational Psychology Dept.
Southern Illinois University at Carbondale

Ann Ramos

Center for English as a Second Language
Southern Illinois University at Carbondale

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E. Maloney & S.M. Tracz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the annual convention of the American Educational Research Association, Montreal, Quebec, Canada, April 13, 1983

TM 830 737

Abstract

Since the numbers of foreign students aspiring to college and graduate degrees in the United States is increasing, there is a need for a reliable and valid English proficiency test which, ideally, would be neither costly nor time consuming. The purpose of this study was twofold. First, the extent to which three tests, the Michigan battery, the Center for English as a Second Language Placement Test, and the Davies Speeded Reading Test of English Language Proficiency (an editing cloze test), tended to measure the same underlying construct was explored. Second, two methods of grading the Davies, the Omission method and the Omission and Commission method, were compared. A total of almost 300 subjects in four groups were tested. Two were homogeneous groups of Malaysians relatively proficient in English. Two consisted of foreign students with diverse language backgrounds. One of these groups had already completed an intensive proficiency course in English, while the other group included students at all levels of English language proficiency. Reliability and validity coefficients were reported along with test means and standard deviations.

In the wake of ever-increasing numbers of foreign students aspiring to degrees in institutions of higher education in the United States, there is growing concern over methods and tools for assessment of English language proficiency.

The Michigan Test of English Language Proficiency and Aural Comprehension and the Test of English as a Foreign Language (TOEFL) are the two most widely administered and accepted predictors of readiness for university and college study of foreign students. The cost of these tests ranges from $19 to $27, and the turn-around time for results varies from two to three weeks. In addition, the test session for the TOEFL lasts approximately three hours, and the Michigan battery two hours. The Davies Speeded Reading Test of English Language Proficiency, on the other hand, takes 25 minutes to administer and grade and costs less than $3.

The present study explores the extent to which the Michigan battery, comprised of the Michigan Test of English Language Proficiency (MTELP) and the Michigan Test of Aural Comprehension (MTAC), the Center for English as a Second Language Placement Test (CESL Placement), and the Davies Speeded Reading Test of English Language Proficiency (Davies) tend to measure the same basic underlying concepts. The concept of proficiency can be viewed as "global mental capacity" (Oller, 1976) which includes listening, speaking, reading, and writing or some subset of those skills.

The existence of voluminous efforts at defining proficiency serve as testimonial to its elusive nature. In order to understand what it

means to be proficient in a language, it is necessary to understand some basic facts about language itself.

Language, says Spolsky, is redundant. Natural communication is generally accompanied by a set of extraneous words, questions or signals which effect reductions in understanding. The following is an example by Spolsky of a cloze test which is predicated on the notion of redundancy in language:

> When for _____, every third _____ is left _____, it is _____ possible to _____. When it is every sixth _____, it is really so simple _____ you can read as if all are there. But _____ second _____ is _____ difficult.

Thus, Spolsky shows that by deleting every second, third, sixth, or nth word, a distorted message results which the native language speaker can interpret. This, language researchers then surmised, is a possible method of measuring language proficiency.

As far back as the 1940s, Bell Telephone Laboratories were conducting research on redundancy in sound signal. "Indications were that the acoustic signal contains twice as much acoustic information as is necessary for a native speaker of the language to understand a telephone message" (Cartier in Jones and Spolsky, 1975). Oller (1972) calls the phenomenon which native speakers seem to possess of knowing what to expect in sequence a "grammar of expectancy." The person who "knows" the language will be able to fill in the blanks and make sense of the message. "Any reduction of redundancy will tend to increase a non-native's difficulty in functioning in a second language more than a native speaker, exaggerating differences and permitting more precise measurement" (Gradman & Spolsky, 1975).

Cloze passages are examples of reduced redundancy. As such, the ability of an individual to produce a suitable response for a cloze bank

is indicative of their proficiency in a given area. Cloze tests were successfully pioneered by Taylor (1953) as a means of measuring readability for native language learners. However, for at least a decade they have been used with both foreign and second language learners. Darnell (1968) obtained a correlation of .83 between a cloze test with a modified scoring system and the TOEFL. Anderson (1971) used cloze procedure to determine readability, hence, appropriateness of materials for English as a Second Language (ESL) classes. Oller and Conrad (1971) administered a cloze test to 100 ESL students. Scores correlated highly with other proficiency measures administered at the same time, especially with a multiple-choice test of general reading proficiency (.80) and a dictation exercise (.82). Alderson (1979) reported a favorable correlation between ESL proficiency and cloze. Numerous other studies attest to the effectiveness of cloze procedure as a viable tool for measuring ESL proficiency.

Cloze tests can be characterized according to their format. Formats include the standard, open-ended cloze, the multiple choice cloze, and the editing cloze. Although other cloze procedures exist, these three formats are the most common. The Davies Speeded Reading Test of English Language Proficiency belongs to the editing cloze category. Typically, such a test consists of a passage into which superfluous words (hereafter called target words) have been inserted. The position of these target words within the passage usually varies from every third to every 18th word and they never appear side by side. Since these superfluous words hinder the meaning of the passage, the examinee must locate these words and mark them out. While text length may vary considerably, the first and last sentences are generally left intact.

There are two types of errors: (a) errors of omission (failure to mark the target words) and errors of commission (erroneously marking a non-target word). Two methods of scoring the editing cloze test also appear in the literature. The first is a two-step method employed by Bowen (1978) and Davies (1975). They compute a raw score by totaling the number of correctly identified target words from which the total number of errors of commission is subtracted. This method, hereafter called the Omission and Commission method, allows for the posibility of negative scores.

The second method of scoring editing cloze tests has been suggested by Mullen (1979). Her method involves totaling correctly identified target words only. Mullen found higher validity when using her method of scoring despite the fact that only one type of error was incorporated. This method is called the Omission method, and negative scores are not possible.

Subjects

Four groups of foreign students were tested in this study. The first two groups (n_1 = 61 and n_2 = 59) consisted of Malaysians whose native language was Malay. These students were graduates of secondary institutions in Malaysia and were selected by their government to be sponsored to study in American universities. Although they were quite proficient when they arrived in the U.S., all attended an intensive English language program before being enrolled in various universities. Their ages ranged from 18 to 24 years, and both sexes were almost equally represented.

The last two groups consisted of foreign students with diverse language backgrounds. The members of these groups were predominantly

males with ages ranging from 18 to 45. Group 3 (n_3 = 173) represents the most heterogeneous combination of students. When tested these students were applying for admission to American institutions, some applied to graduate school, others to baccalaureate programs, and still others to English as a Second Language programs. Obviously, the degree of English proficiency of this group varied greatly. Group 4 (n_4 = 44) students already had completed an intensive English course prior to testing. Completion of this course, however, does not guarantee proficiency in English.

Instrument Description

The Michigan Battery has two components. The first is the Michigan Test of English Language Proficiency (MTELP), which consists of 45 grammar items, 40 vocabulary items, and 20 reading items. The second is the Michigan Test of Aural Comprehension (MTAC). Here students are required to listen to a 30-minute tape and answer questions in a test booklet during pauses in the tape. Although the Michigan battery has two components, results were reported in terms of composite scores only. Different forms of the Michigan were used in different testing sessions.

The CESL Placement Test, as its name implies, was an instrument designed to place students in one of five levels of English language classes prior to their entry into regular American universities. The English language proficiency of these students ranges from the rank beginner (Level 1) to those capable of graduate work in the U.S. (Level 5). Like the Michigan battery the CESL Placement is a composite test consisting of subtests in listening comprehension, grammar and reading. Again results were reported for the composite scores. Different forms of the CESL Placement were used in different testing sessions.

The Davies Speeded Reading Test of English Language Proficiency (Davies) was constructed in 1967 by Alan Davies. The Davies is an editing cloze task with 1,186 words and 196 target words. Paragraph one intersperses non-English words as target words, while the remainder of the examination uses English words as target words. This test is speeded and examinees are given ten minutes to complete it. The Davies was graded in two ways (errors of Omission only and errors of Omission and Commission) so a comparison between methods could be made.

Testing Procedures

Prior to any instructionin the U.S., students in groups 1 and 2 were randomly assigned to one of the two groups. Students in group 1 were tested with the CESL Placement Test, followed by the Michigan battery, and then the Davies. Students in group 2 were tested with the Michigan battery, the CESL Placement, and the Davies. Thus any systematic bias due to the order in which the tests were administered was avoided. Pretesting and posttesting followed the same pattern.

Only one testing session could be arranged for the subjects in Groups 3 and 4 since many were enroute to other areas of the country. Students in group 3 were tested in this study prior to any instruction in the U.S. They were given the CESL Placement first and then the Davies. Students in group 4 were given the Michigan battery and then the Davies after attending an intensive English language program.

Results

The means, standard variations, and ranges for all variables for Groups 1, 2, 1 and 2 combined, 3, 4, and 1 through 4 combined appear in Tables 1 through 6 respectively. For the Malaysian groups (Tables 1,

2, and 3) where two testing sessions were possible, all means were higher at posttesting than at pretesting. This was to be expected since a 10-week period of English instruction intervened. The means for the two groups of heterogeneous language backgrounds (Tables 4 and 5) are much lower than the means for the Malaysians. Since many of these students (unlike the Malaysians) had only limited exposure to English before their arrival in the U.S., these are also expected results.

The means for all groups combined are reported in Table 6. It should be noted that Group 4 was only tested once with the Michigan, Group 3 was only tested once with the CESL Placement, and both Groups 3 and 4were only tested once with the Davies. Here the Michigan posttest scores for Malaysians were combined with the Michigan scores for students in Group 4 since the latter were tested after attending the English training program. The CESL placement pretest scores for Malaysians were combined with the only CESL placement scores for students in Group 3 since the latter had had no formal English training in the U.S. at the time of testing. Malaysian posttest scores for the Davies were combined with the Davies for Groups 3 and 4. Both grading methods were reported.

Two types of reliabilities were reported for Groups 1 and 2 combined. They appear in Table 7. The delayed alternate-form reliabilities for the two criterion measures, the Michigan and the CESL Placement, are .89 and .85 respectively. Test-retest reliabilities for the Davies using both grading methods, first Omission and then Omission and Commission, were .64 and .68 respectively. Unfortunately, only one testing period could be arranged for Groups 3 and 4 so neither delayed alternate-form nor test-retest reliabilities could be calculated.

The validity analysis or the results of the intercorrelations of all instruments and grading methods for pretesting, posttesting, and one testing sessions appear in Tables 7 through 10. Correlations between the Michigan and the CESL Placement ranged from .84 to .87 in Table 7, and it was .86 in Table 10.

The validity correlations between the Michigan and the Davies Omission method ranged from .61 to .68 in Table 7 for the Malaysian groups. A similar correlation between the same variables was .55 in Table 9 for the heterogeneous group and .70 in Table 10 for the combined groups. Correlations between the Michigan and the Davies Omission and Commission method ranged from .66 to .72 in Table 7 for the Malaysians. A similar correlation between the same variables was .53 in Table 9 for those of diverse language backgrounds and .72 in Table 10 for combined groups. So the validity coefficients were generally higher when using the Omission and Comission method.

Correlations between the CESL Placement and the Davies Omission method ranged from .61 to .74 in Table 7 for Malaysians. The correlation between the same variables was .70 in Table 8 for the heterogeneous group and .73 in Table 10 for the combined groups. The validity coefficients between the CESL placement and the Davies Omission and Commission method ranged from .66 to .75 in Table 7 for Malaysians. They were .71 in Table 8 for the heterogeneous group and .74 in Table 10 for the combined groups for the same variables. Again, the validity coefficients were higher using the Omission and Commission method of grading the Davies.

The correlation coefficients for the Davies between the two methods of grading, the Omission method with the Omission and Commission method,

for Malaysian were .95 for the pretest and .98 for the posttest (Table 7). These correlations were .99 for the heterogeneous group with no U.S. English language training (Table 8) and .99 for the heterogeneous group with U.S. English language training (Table 9). The correlation between the Davies Omission method and the Omission and Commission method for all subjects was .99 (Table 10).

Discussion

The reliabilities for the Davies were not exceptionally high. Although it is possible that the reliabilities would be higher with less homogeneous groups than those consisting only of Malaysians with a good background in English, this may not necessarily be the case. Further research needs to be done with the Davies within various language families.

The high correlations between the Michigan and the CESL Placement indicate that these tests were measuring the same construct. However, the correlations between the criterion measures and the Davies were all considerably lower. Also, the mean values for the four groups seemed to indicate that the Davies is potentially a better proficiency instrument than a diagnostic one.

When comparing the two methods for grading the Davies, it was found that the means for the Omission grading method were higher than the means for the Omission and Commission grading method in all cases. This was to be expected since the sum of a different type of error was subtracted from the total number of correct target words with the Omission and Commission method.

Also, the correlations between the criterion measures and the Davies graded with the Omission and Commission method were higher in all cases except one. A possible reason for these results is that correlations are attenuated by a restricted range. Since the scores can assume positive and negative values with the Omission and Commission method, the range is extended. The restricted range of the Omission method tends to depress the correlations. It is helpful to view the problem of restricted range from a conceptual viewpoint. The Omission method of grading only gave information about one type of error, the inability to identify target words. The Omission and Commission method gave information about two types of errors so that every word in the test was important.One could conclude, then, that the Omission and Commission method of grading was superior.

Table 1

Means, Standard Deviations, and Ranges
for the Michigan, CESL Placement, the Davies (Omission),
and the Davies (Omission and Commission) of Group 1
(n = 61)

	Mean	Standard Deviation	Range	
			Low	High
Michigan Pretest	79.65	11.33	36	93
Michigan Posttest	82.55	7.92	57	95
CESL Placement Pretest	73.03	12.20	38	91
CESL Placement Posttest	78.30	7.77	52	90
Davies Pretest (Omission)	85.64	27.33	11	136
Davies Posttest (Omission)	101.02	25.30	38	138
Davies Pretest (Omission and Commission)	79.70	29.20	-22	132
Davies Posttest (Omission and Commission)	90.77	26.18	38	134

Table 2

Means, Standard Deviations, and Ranges
for the Michigan, CESL Placement, the Davies (Omission),
and the Davies (Omission and Commission) of Malaysian Group 2
(n = 59)

	Mean	Standard Deviation	Range	
			Low	High
Michigan Pretest	79.90	9.44	56	94
Michigan Posttest	82.98	6.40	70	95
CESL Placement Pretest	75.97	9.09	55	90
CESL Placement Posttest	77.18	9.09	59	92
Davies Pretest (Omission)	72.08	21.74	36	123
Davies Posttest (Omission)	107.30	41.21	1	189
Davies Pretest (Omission and Commission)	68.49	20.72	34	119
Davies Posttest (Omission and Commission)	97.66	39.26	1	186

Table 3

Means, Standard Deviations, and Ranges
for the Michigan, CESL Placement, the Davies (Omission),
and the Davies (Omission and Commission) for Combined Malaysian
Groups 1 and 2 (n = 120)

	Mean	Standard Deviations	Ranges Low	Ranges High
Michigan Pretest	79.78	10.38	36	94
Michigan Posttest	82.76	7.18	57	95
CESL Placement Pretest	74.49	10.83	38	91
CESL Placement Posttest	77.74	8.44	52	92
Davies Pretest (Omission)	78.98	25.56	11	136
Davies Posttest (Omission)	104.13	34.51	1	189
Davies Pretest (Omission and Commission)	72.67	25.62	-22	132
Davies Posttest (Omission and Commission)	94.16	33.30	1	186

Table 4

Means, Standard Deviations, and Ranges for the CESL Placement, the Davies (Omission), and the Davies (Omission and Commission) for Heterogeneous Nationalities Group 3 (n = 173)

	Mean	Standard Deviation	Range	
			Low	High
CESL Placement	57.59	18.90	4	90
Davies (Omission)	51.32	29.83	0	151
Davies (Omission and Commission)	46.46	27.89	0	147

Table 5

Means, Standard Deviations, and Ranges for the Michigan, the Davies (Omission), and the Davies (Omission and Commission) for Heterogeneous Group 4 (n = 44)

	Mean	Standard Deviation	Range	
			Low	High
Michigan	74.73	7.40	58	86
Davies (Omission)	48.91	22.22	0	96
Davies (Omission and Commission)	45.05	20.40	0	88

Table 6

Means, Standard Deviation, Ranges, and Sample Sizes for the Michigan, the CESL Placement, the Davies (Omission), and the Davies (Omission and Commission) for All Nationalities Combined, Groups 1, 2, 3, and 4

	Mean	Standard Deviation	Range Low	Range High	Sample Size
Michigan	80.73	8.02	57	95	164
CESL Placement	64.59	18.07	4	91	293
Davies (Omission)	69.91	39.84	0	189	337
Davies (Omission and Commission)	63.26	37.06	0	186	337

Table 7

Intertest and Test-Retest Correlations of the Michigan, CESL Placement, Davies (Omission), and Davies (Omission and Commission) for Combined Malaysian Groups 1 and 2 (n = 120)

	1	2	3	4	5	6	7
1. Michigan Pretest							
2. Michigan Posttest	.89						
3. CESL Placement Pretest	.88	.86					
4. CESL Placement Posttest	.87	.84	.85				
5. Davies (Omission) Pretest	.68	.67	.61	.74			
6. Davies (Omission) Posttest	.61	.62	.62	.63	.64		
7. Davies (Omission and Commission) Pretest	.69	.72	.70	.75	.95	.64	
8. Davies (Omission and Commission) Posttest	.66	.67	.66	.67	.66	.98	.68

Table 8

Intertest Correlations for the CESL Placement, the Davies (Omission), and the Davies (Omission and Commission) for Heterogeneous Nationalities Group 3 (n = 173)

	1	2
1. CESL Placement		
2. Davies (Omission)	.70	
3. Davies (Omission and Commission)	.71	.99

Table 9

Intetest Correlations for the Michigan, the Davies (Omission), and the Davies (Omission and Commission) for Heterogeneous Nationalities Group 4 (n = 44)

	1	2
1. Michigan		
2. Davies (Omission)	.55	
3. Davies (Omission and Commission)	.53	.99

Table 10

Intertest Correlation and Sample Sizes for the Michigan, the CESL Placement, the Davies (Omission), and the Davies (Omission and Commission) for all Nationalities Combined, Groups 1, 2, 3, and 4

	1	2	3
1. Michigan			
2. CESL Placement	.86 (120)		
3. Davies (Omission)	.70 (164)	.73 (293)	
4. Davies (Omission and Commission)	.72 (164)	.74 (293)	.99 (337)

Bibliography

Alderson, J. Charles. The cloze procedure and proficiency in English as a foreign language. TESOL Quarterly, 1979b, 13, 219-227.

Allen, M.J., & Yen, W.M. Introduction to Measurement Theory. Monterey, California: Brooks/Cole Pub. Co., 1979.

Anderson, Jonathan. Selecting a suitable "reader:" Procedures for teachers to assess language difficulty. RELC Journal, 1971, 2, 35-42.

Bowen, J.D. The identification of irrelevant distraction: An editing task. TESL Reporter, 1978, 12 (1), 1-3; 14-16.

Darnell, K. The development of English language proficiency tests of foreign students, using a clozentropy procedure. Boulder, Colorado: University of Colorado, 1968. (ERIC Document Reproduction Service No. ED. 024 039).

Davies, Alan. Two tests of speeded reading. In Randall L. Jones & Bernard Spolsky (eds.), Testing language proficiency. Arlington, Virginia: Center for Applied Linguistics, 1975.

Gradman, Harry L. & Spolsky, Bernard. Reduced redundancy testing: A progress report. In Randall L. Jones & Bernard Spolsky (eds.), Testing language proficiency. Arlington, Virginia: Center for Applied Linguistics, 1975.

Mehrens, W.A., & Lehmann, I.J. Measurement and Evaluation in Education and Psychology (2nd ed.). New York: Holt, Reinhart and Winston, 1975.

Mullen, Karen A. An alternative to the cloze test. In C.A. Yorio, K. Perkins, & J. Schachter (eds.), On TESOL 1979: The Learner in focus. Washington, D.C.: Teachers of English to Speakers of

Oller, John W., Jr. Scoring methods and difficulty levels for cloze tests of ESL proficiency. Modern Language Journal, 1972, 56, 151-158.

Oller, John W., Jr. A program for language testing research. Language Learning, 1976, Special Issue Number 4, 141-166.

Oller, John W., Jr. & Conrad, C. The cloze procedure and ESL proficiency. Language Learning, 1971, 21, 183-196.

Taylor, Wilson L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 414-438.